Europäisches Patentamt | European Patent Office | Office européen des brevets

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

99480017.5

Der Präsident des Europäischen Patentamts:
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

Arlette Fiedler

A. Fiedler

DEN HAAG, DEN
THE HAGUE, 04/05/99
LA HAYE, LE

Europäisches Patentamt
European Patent Office
Office européen des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.: 99480017.5
Demande n°:

Anmeldetag:
Date of filing: 30/03/99
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
INTERNATIONAL BUSINESS MACHINES CORPORATION
Armonk, NY 10504
UNITED STATES OF AMERICA

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
Multiple ARP functionality for an IP data transmission system

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat: State: Pays:	Tag: Date: Date:	Aktenzeichen: File no. Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:
/

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE
Etats contractants désignés lors du dépôt:

Bemerkungen:
Remarks:
Remarques:

Multiple ARP functionality for an IP data transmission system

Technical field

The present invention deals with a new way for load balancing outgoing IP packets from an IP host such as a large Web server, and relates in particular to a multiple ARP functionality for an IP data transmission system.

Background

Modern digital networks are made to operate over different transmission media and interconnect upon request a very large number of users (e.g. hosts) and applications through fairly complex digital communication networks.

Due to the large variety of users' profiles and distributed applications, the traffic is becoming more and more bandwidth consuming, non-deterministic and requiring more connectivity. This has been the driver for the emergence of fast packet switching techniques in which data from different origins are chopped into fixed or variable length packets or datagrams, and then transferred, over high speed digital networks, between a data source and a target terminal equipment.

Several types of networks have been installed throughout the world, which need to be interconnected (e.g. via so called Routers) to optimize the possibilities of organizing traffic between source hosts and target hosts located anywhere in the world. This is made possible by using so-called internetworking.

Internetwork (also referred to as internet) facilities use a set of networking protocols such as Transmission Control Protocol/Internet Protocol (TCP/IP) developed to allow cooperating host computers to share resources across the internetwork. TCP/IP is a set of data communication protocols that are

referred to as internet protocol (IP) suite. Because TCP and IP are the best known, it has become common to use the term TCP/IP to refer to the whole protocol family. TCP and IP are two of the protocols in this suite. Other protocols of the suite are User Datagram Protocol (UDP), Address Resolution Protocol (ARP), Real Time Protocol (RTP) etc...

An internet may thus be a collection of heterogeneous and independent networks using TCP/IP, and connected together by routers. The administrative responsibilities for an internet (e.g. to assign IP addresses and domain names) can be within a single network (LAN) or distributed among multiple networks.

When a communication of data has to be established from a source host to a particular IP destination over an IP network, there is a number of methods to determine the first hop router of the network towards this destination. These include running (or snooping) dynamic routing protocol such as Routing Information Protocol (RIP) or Open Shortest Path First (OSPF) version, running an ICMP router discovery client or using a statically configured default route.

Running a dynamic routing protocol on every end-host may be infeasible for a number of reasons, including administrative overhead, processing overhead, security issues, or lack of a protocol implementation for some platforms. Neighbor or router discovery protocols may require active participation by all hosts on a network, leading to large timer values to reduce protocol overhead in face of large numbers of hosts. This can result in a significant delay in the detection of a lost (i.e., dead) neighbor, which may introduce unacceptably long "black hole" periods.

The use of a statically configured default route is quite popular, it minimizes configuration and processing overhead on the end-host and is supported by virtually every IP implementation. This mode of operation is likely to persist as Dynamic Host Configuration Protocols (DHCP) are deployed, which typically provide configuration for an end-host IP address and

default gateway. However, this creates a single point of failure. Loss of the default router results in a catastrophic event, isolating all end-hosts that are unable to detect any alternate path that may be available.

One solution to solve this problem is to allow hosts to appear to use a single router and to maintain connectivity even if the actual first hop router they are using fails. Multiple routers participate in this protocol and in concert create the illusion of a single virtual router. The protocol ensures that one and only one of the routers is forwarding packets on behalf of the virtual router. End hosts forward their packets to the virtual router. The router forwarding packets is known as the active router. A standby router is selected to replace the active router should it fail. The protocol provides a mechanism for determining active and standby routers, using the IP addresses on the participating routers. If an active router fails, a standby router can take over without a major interruption in the host's connectivity.

Another similar approach is the use of Virtual Router Redundancy Protocol (VRRP) designed to eliminate the single point of failure inherent in the static default routed environment. VRRP specifies an election protocol that dynamically assigns responsibility for a virtual router to one of the VRRP routers on a LAN. The VRRP router controlling the IP address(es) associated with a virtual router is called the Master, and forwards packets sent to these IP addresses. The election process provides dynamic fail-over in the forwarding responsibility should the Master become unavailable. Any of the virtual router's IP addresses on a LAN can then be used as the default first hop router by end-hosts. The advantage gained from using VRRP is a higher availability default path without requiring configuration of dynamic routing or router discovery protocols on every end-host.

Unfortunately the two above solutions cannot provide load balancing for a given host's traffic because only the router that answered the ARP is used. Also, customers are reluctant

to change their main router configuration to enable such a function.

Summary of the invention

Accordingly, the object of the invention is to provide a data transmission system including an IP network wherein it is the IP host which selects directly the default router thereby improving load balancing and high availability.

Another object of the invention is to enable an IP source host to be aware of the availability of a set of candidate default routers and to select one of them dynamically, ensuring both load balancing and high availability.

Another object of the invention is a method of selecting a router amongst a set of routers for an IP host in a data transmission system including an IP network.

Therefore, the invention relates to a data transmission system for transmitting packetized data from an IP host having at least an IP layer and a network layer to a plurality of work-stations by the intermediary of an IP network and wherein the IP host is connected to the IP network via a layer 2 network interfacing the IP network by a set of routers, the IP host further including a Multiple Address Resolution Protocol (MARP) layer between the IP layer and the network layer for selecting one of the set of routers in response to the next hop IP address provided by the IP layer to the multiple ARP layer when a packet of data is to be transmitted from the IP host to one of the workstations.

Brief description of the drawings

The above and other objects, features and advantages of the invention will be better understood by reading the following more particular description of the invention in conjunction with the accompanying drawings wherein :

Fig. 1 represents schematically a data transmission system wherein an IP host can select one router amongst a set of routers according to the invention.

Fig. 2A and 2B represent respectively the MARP table and the ARP table used in combination to achieve the method according to the invention.

Fig. 3 is a flow chart of the method of selecting a router according to the invention.

Detailed description of the invention

In reference to Fig. 1, the invention is implemented in a data transmission system wherein an IP host has to transmit data to one or several workstations 12, 14 via an IP network 16 such as Internet. It can be assumed that IP host 10 is connected to IP network 16 by means of a layer 2 network such as Local Area Network (LAN) 18 which is interfacing IP network 16 by a set of input routers 20, 22 and 24. The IP packets are routed over the IP network via a plurality of routers (not shown) until an output router 26 connected directly (or by means of a layer 2 network) to workstations 12 or 14.

As illustrated in Fig. 1, to communicate over the IP network, IP host 10 must implement a layered set of protocols 28 referred as the Internet protocol suite. Without the invention the protocol suite would be used as follows :

- the application layer 30 (level 5) generates a data stream to be sent and passes this data stream to a transport layer,
- the transport layer (level 4) such as TCP layer 32, segments the data stream into packets and passes the packet to the IP layer for routing to the destination IP address with an added TCP Header,
- the IP layer 34 finds the next hop IP address based upon the destination IP address. Normally, with the IP Host

which does not run a routing protocol, this address is a default entry that leads to a default router.

- IP layer 34 passes the IP packet to the network layer (not shown) with an added IP header information. As a side parameter, the IP layer informs the network layer of the next hop IP address.
- the network layer resolves the next hop IP address into a network address of the default router using the ARP protocol and transmits the packet over the IP network.

The invention introduces a new layer between IP layer 34 and the network layer, a Multiple ARP (MRAP) layer 36. Therefore, IP layer 34 passes the packet and the next hop IP address to MARP layer 36 instead of the network layer. As explained below, this MARP layer runs an algorithm to determine the best physical router 20, 22 or 24 based on parameters defined in the packet such as source and destination addresses and ports.

At the destination workstation 12, a reciprocal protocol suite 38 is implemented. Namely, the network layer passes the IP packets to IP layer 40 which transfers the packets to TCP layer 42 for reassembling them into a data stream communicated to the application layer 44. Note that workstation 12 does not include a MARP layer since such a layer is not required for receiving data, but could also be an IP host provided with a MARP layer used to transmit IP packets over the network in the same way as made by IP host 10.

The MARP layer operates with a table called the MARP table represented in Fig. 2A. The MARP table maps the next hop IP address into a set of candidate IP addresses corresponding to candidate routers amongst the set of routers 20, 22 and 24 interfacing the IP network as illustrated in Fig. 1. In the simplest form, there is only one entry in the MARP table for the default router, that points on the set of candidate routers which can act as default routers.

The candidate routers associated with the IP addresses in MARP table can either be configured to the MARP layer via a configuration tool, or be dynamically acquired using a learning protocol such as an extension to the Dynamic Host configuration Protocol (DHCP).

As some ones of the candidate routers may not be active at a given time, the MARP layer uses the ARP table provided by the network layer as illustrated in Fig. 2B. The ARP table maps the IP addresses provided by the MARP table into network addresses.

Referring now to Fig. 3, the selection of an active router is as follows. When an IP packet is to transmit over the network, the MARP layer is called by the IP Layer and the next hop IP address (usually that of the default router) is provided as a parameter for looking up the MARP table (step 50). If the next hop IP address matches an entry in the MARP table (step 52), an associated list of candidate routers is built (step 54). The candidate routers are then checked in the ARP table, one by one (step 56). A determination is made (step 58) of the candidate routers which have a recent entry in the ARP table, and these routers are selected as active candidate routers. Note that, if no active candidate routers can be determined (step 58), the packet is destroyed (step 60).

Out of the list of the active candidate routers, the MARP layer selects (step 62) one IP address corresponding to a candidate router that is passed to the network layer as a substitute of the original next hop IP address as selected by the IP layer. In the preferred embodiment, this selection is performed on a per packet basis, without an history of previous selection, but this is not the only possible selection algorithm. Other techniques like round robin or byte wise weighting mechanisms could be used alternatively. The preferred implementation uses an hash coding technique as described in European Patent Application n° 98480062.3, in order to stick a TCP connection to a same candidate router as long as the candidate topology is left unchanged. The hash coding uses the

destination IP address and the pair of ports in the packets. These are mingled with the candidate routers IP addresses, one by one. The highest resulting hash value is selected. Weight coefficients may be used to modify the statistical expectancy of each individual candidate, in order to match their capacity.

At last, the IP packet is sent to the network layer (step 64) for it to be transmitted to the candidate router which has been selected. It must be noted that the IP packet will directly be sent to the network layer when no match has been found (step 52) in looking up the MARP table because the next hop IP address corresponds to a router or a host which is not required to be substituted.

It must be noted that the MARP layer only uses candidates that are already present in the ARP table. As a consequence, MARP layer uses an out-of-band technique to be sure that the ARP table is correctly filled with all the up-to-date information. In the preferred embodiment, periodic void packets like ICMP echo are transmitted to the non-active routers, that is candidate routers which are not present in the ARP table. Upon such packets, the ARP function in the network layer will automatically refresh the entry by using the ARP protocol. Also, at the initial time, one such packet is sent to all the configured routers to preset the ARP table before a single data is issued by an application layer.

The ARP function ensures the freshness of the ARP table by aging the entries and flushing the older ones. To maintain the status of active candidate routers, the preferred method consists in resetting the age of an entry each time a packet is received from a matching network address. Also, if an entry gets old, but before it is flushed by ARP, MARP may flush the ARP table entry right before it passes a packet to the Network layer with the next hop IP address pointing on that router. Again, this forces the Network layer to use ARP procedures to check for the router availability.

Claims

1. Data transmission system for transmitting packetized data from an IP host (10) having at least an IP layer (34) and a network layer to a plurality of workstations (12, 14) by the intermediary of an IP network (16) and wherein said IP host is connected to said IP network via a layer 2 network (18) interfacing said IP network by a set of routers (20, 22, 24) ;

said system being characterized in that said IP host further includes a Multiple Address Resolution Protocol (MARP) layer (36) between said IP layer and said network layer for selecting one of said set of routers in response to the next hop IP address provided by said IP layer to said multiple ARP layer when a packet of data is to be transmitted from said IP host to one of said workstations.

2. Data transmission system according to claim 1, wherein said IP host (10) is provided with an Address Resolution Protocol (ARP) in charge of resolving any IP address into a network address of the router to be used in said layer 2 network (18) by mapping in an ARP table said IP address into the network address of an active router amongst said set of routers (20, 22, 24).

3. Data transmission system according to claim 2, wherein said MARP layer (36) includes a MARP table mapping said next hop IP address into a list of candidate routers amongst said set of routers (20, 22, 24), said candidate routers being mapped in said ARP table into active candidate routers able to be used as routers for transmitting said packet of data from said IP host (10) to one of said workstations (12, 14).

4. Data transmission system according to claim 3, wherein one router is selected amongst said active candidate routers by using a hash coding method based upon the destination IP address, the pair of source and destination

ports in said packet of data to be transmitted, and the active candidate router IP addresses.

5. Method of selecting a router by an IP host (10) in a data transmission system transmitting packetized data from said IP host having at least an IP layer (34) and a network layer to a plurality of workstations (12, 14) by the intermediary of an IP network (16) and wherein said IP host is connected to said IP network via a layer 2 network (18) interfacing said IP network by a set of routers (20, 22, 24) ;

said method being characterized by determining a list of candidate routers amongst said set of routers and determining a list of active candidate routers amongst said candidate routers before selecting said router to be used for transmitting said packet of data amongst said list of active candidate routers.

6. Method according to claim 5, wherein said step of determining said list of active candidate routers is performed by a Multiple Address Resolution Protocol (MARP) layer (36) between the IP layer (34) and the network layer of said IP host (10).

7. Method according to claim 6, wherein said step of determining said list of candidate routers is performed by said MARP layer (36) by a look up of a MARP table using the next hop IP address as entry.

8. Method according to claim 7, wherein said step of selecting said router to be used for transmitting said packet of data is performed by using a hash coding technique based upon the destination IP address, the pair of source and destination ports in said packet of data to be transmitted, and the active candidate router IP addresses.

Multiple ARP functionality for an IP data transmission system

Abstract

Data transmission system for transmitting packetized data from an IP host (10)having at least an IP layer (34) and a network layer to a plurality of workstations (12, 14) by the intermediary of an IP network (16) and wherein the IP host is connected to the IP network via a layer 2 network (18) interfacing the IP network by a set of routers (20, 22, 24). The IP host further includes a Multiple Address Resolution Protocol (MARP) layer (36) between the IP layer and the network layer for selecting one of the set of routers in response to the next hop IP address provided by the IP layer to the multiple ARP layer when a packet of data is to be transmitted from the IP host to one of the workstations.

Fig. 1

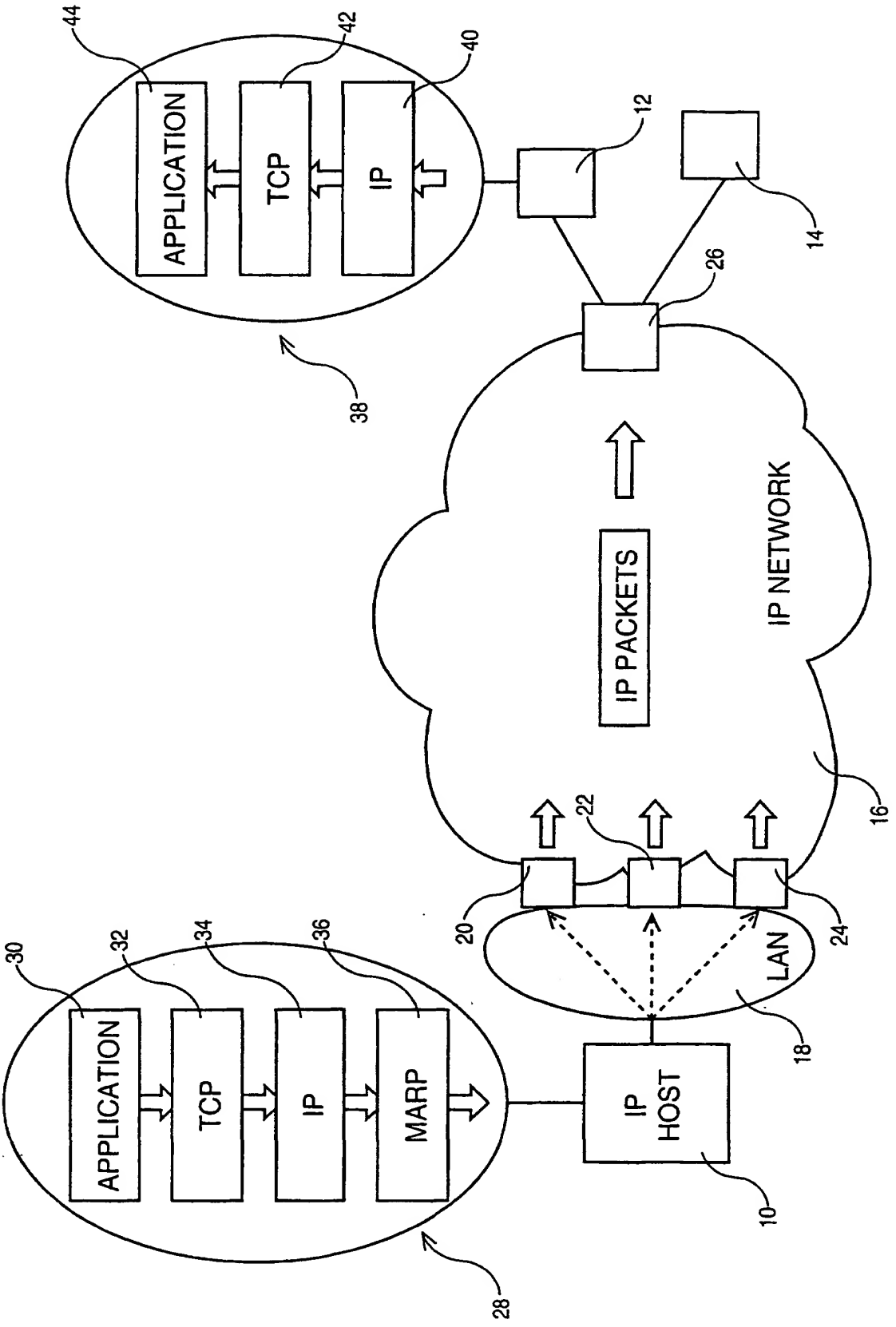

FIG. 1

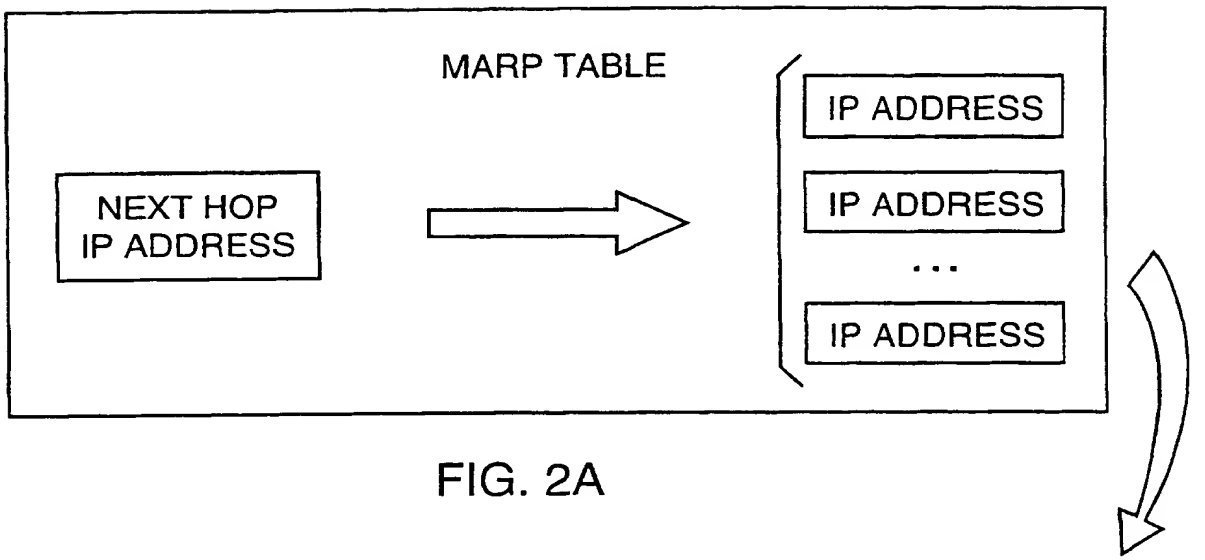

FIG. 2A

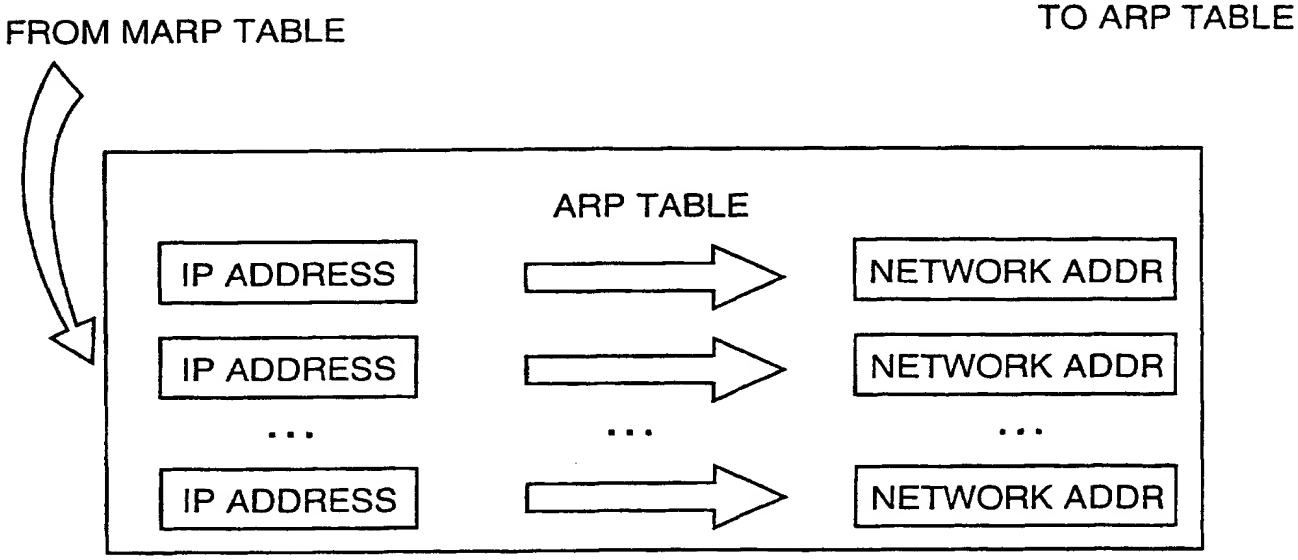

FIG. 2B

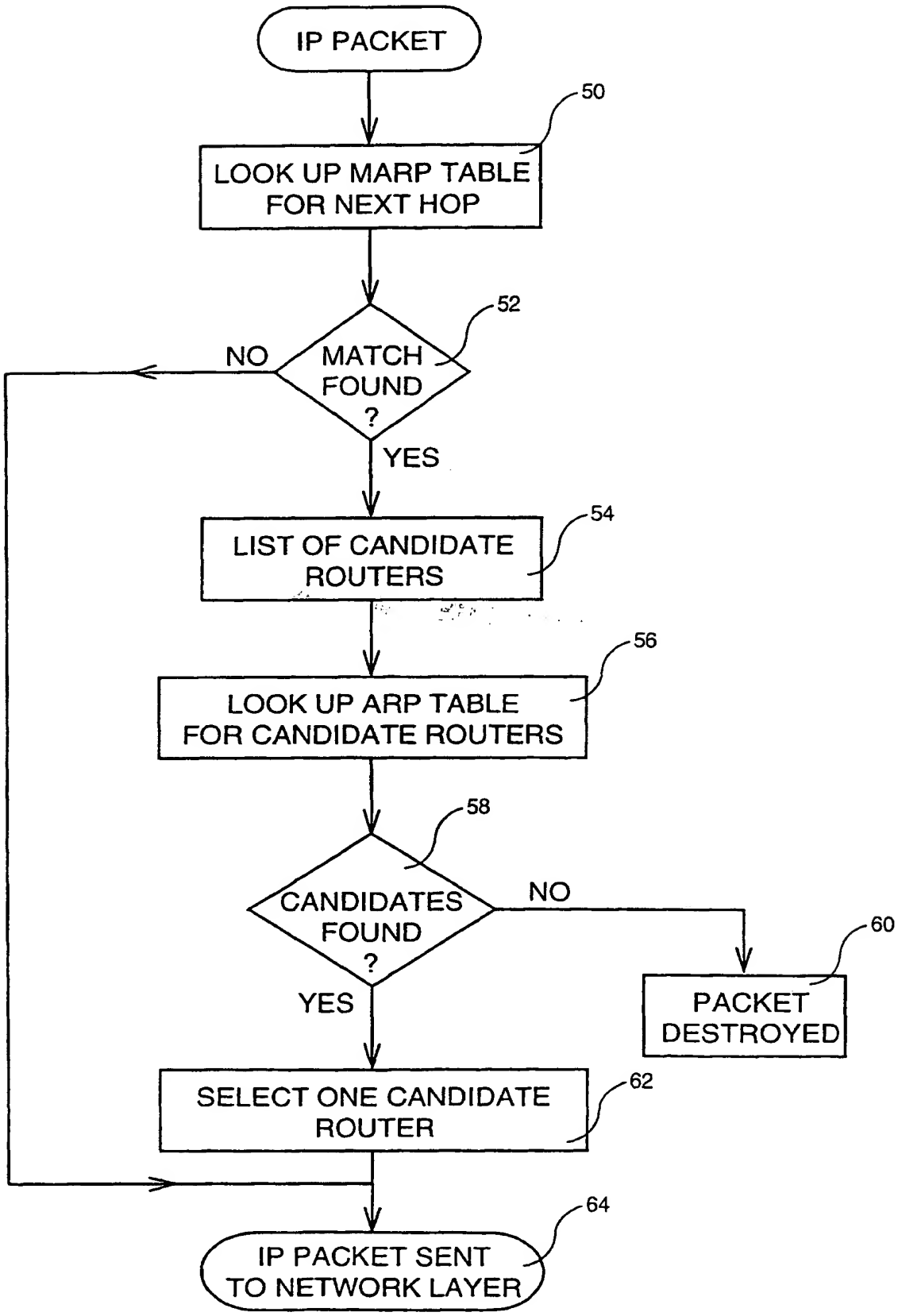

FIG. 3